



Guidelines to use Transfer Learning for Motor Imagery Detection: an experimental study

Laurent Bougrain, Sébastien Rimbert, Pedro Luiz Coelho Rodrigues, Geoffrey Canron, Fabien Lotte

► To cite this version:

Laurent Bougrain, Sébastien Rimbert, Pedro Luiz Coelho Rodrigues, Geoffrey Canron, Fabien Lotte. Guidelines to use Transfer Learning for Motor Imagery Detection: an experimental study. NER 2021 - 10th International IEEE/EMBS Conference on Neural Engineering, May 2021, Virtual, United States. hal-03147055

HAL Id: hal-03147055

<https://inria.hal.science/hal-03147055>

Submitted on 19 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Guidelines to use Transfer Learning for Motor Imagery Detection: an experimental study

Laurent Bougrain¹, Sébastien Rimbart², Pedro L. C. Rodrigues³, Geoffrey Canon¹ and Fabien Lotte⁴

Abstract—Brain-Computer Interfaces (BCI) based on Motor Imagery (MI) shown promising results for motor recovery, intraoperative awareness detection or assistive technology control. However, they suffer from several limitations due to the high variability of electroencephalographic (EEG) signals, mainly lengthy and tedious calibration times usually required for each new day of use, and a lack of reliability for all users. Such problems can be addressed, to some extent, using transfer learning algorithms. However, the performance of such algorithms has been very variable so far, and when they can be safely used is still unclear. Therefore, in this article, we study the performance of various state-of-the-art Riemannian transfer learning algorithms on a MI-BCI database (30 users), for various conditions: 1) supervised and unsupervised transfer learning; 2) for various amount of available training EEG data for the target domain; 3) intra-session or inter-session transfer; 4) for both users with good and less good MI-BCI performances. From such experiments, we derive guidelines about when to use which algorithm. Re-centering the target data is effective as soon as a few samples of this target set are taken into account. This is true even for an intra-session transfer learning. Likewise, re-centering is particularly useful for subjects who have difficulty producing stable motor imagery from session to session.

I. INTRODUCTION

Brain-Computer Interfaces (BCI) are communication and control systems that enable their users to send commands to a computer using only their brain activity, usually measured with electroencephalography (EEG) [1]. One of the most prominent BCI types of interaction are Motor Imagery (MI)-based BCI, in which users control a system by performing MI tasks, e.g., imagining hand or foot movements, that can be detected from EEG signals. MI-BCI are very promising for numerous applications, e.g., to control assistive technologies, for intraoperative awareness detection, for entertainment or for post-stroke motor rehabilitation [1], [2].

However, MI-BCIs suffer from several limitations, two serious ones being 1) their lack of reliability for many practical applications [3]; 2) their lengthy and tedious calibration times, generally required for each new day of use [4]. To address their limited reliability, various advanced machine

learning algorithms were developed to robustly decode MI-related EEG signals [3]. Among them, Riemannian classifiers, which process EEG signals represented as covariance matrices, are arguably the current state-of-the-art methods in the field [5], [3], [6]. In order to address the lengthy calibration times, transfer learning methods were developed to transfer classifiers and/or features from a source domain, e.g., data from past BCI users or past BCI sessions of the user, to a target domain, e.g., data from a new BCI user or a new session of the user. This has enabled to reduce or even suppress the need for calibration for the target domain [4], [3]. More recently, algorithms providing the best of both worlds were proposed, i.e., Riemannian transfer learning algorithms [7], [8]. Such approaches proved useful in practice, and led to improved classification performances and/or reduced calibration times on average. However, their effects were also very variable across users and contexts. This prompted the community to raise the following question: “When do transfer learning approaches work in BCI?” [9].

This paper contributes answers to this question through an experimental study. In particular, we study the performance of various state-of-the-art Riemannian transfer learning algorithms on a relatively large MI-BCI data base (30 users), for various conditions: 1) supervised and unsupervised transfer learnings; 2) for various amount of available training EEG data for the target domain; 3) intra-session or inter-session transfer; 4) for both users with good and less good MI-BCI performances. From such experiments, we derive guidelines about when to use which algorithm.

II. MATERIAL AND METHODS

A. Participants

30 right-handed healthy subjects (12 females; aged 23.8 years old; STD = 8.4) were recruited for this study. The subjects had no medical history that could have influenced the task (i.e., diabetes, antidepressant treatment, or neurological disorders). All participants signed an informed consent approved by the local ethical committee of Inria (COERLE, approval number: 2016-011/01).

B. Experimental procedure

Each participant took part in one experiment of 70 min divided into 4 steps: (1) installation of the EEG cap (20 min); (2) a first session of Kinesthetic MI (KMI) during which participants had to perform one specific right-hand KMI task of grasping (including a calibration and a testing steps; 15 min); (3) a second session of KMI (including a calibration and a testing steps; 15 min); (4) uninstallation and debriefing

¹Laurent Bougrain and Geoffrey Canon are with dept. of complex system, Artificial intelligence and robotics at Loria, University of Lorraine, France laurent.bougrain@loria.fr

²Sébastien Rimbart is with the PErSEUs team, Université de Lorraine, UR 7312, Metz, F-57045, France sebastien.rimbart@inria.fr

³Pedro L. C. Rodrigues is with the Parietal team at Inria Saclay, 91120 Palaiseau, France pedro.rodrigues@inria.fr

⁴Fabien Lotte is with Inria Bordeaux Sud-Ouest / LaBRI (CNRS / Univ. Bordeaux / Bordeaux INP), 33405, Talence, France fabien.lotte@inria.fr

(20 min). During both sessions 1 and 2, all subjects were seated in front of a screen showing the non-immersive virtual environment of the *Grasp-IT* BCI (for more details, see [10]), which is composed of a three-color traffic light and a virtual right hand. During BCI use, the calibration step enabled the collection of EEG data from both classes (right hand KMI and resting state) and the testing step provided opportunity for the participant to interact with the Grasp-IT interface. The whole calibration step consisted of one run with 40 trials. During each trial, users were invited to perform the KMI of grasping continuously during 4 seconds, as soon as the light turned green and while it remained so. The rest condition was similarly indicated by the red light, lasting 6 seconds. Then, the orange light along with the red one, lasting 2 seconds, warned the subject that the KMI would start soon to prevent motor preparation by anticipation by the end of the red light period.

C. Electrophysiological recordings

EEG signals were recorded using the OpenViBE software platform with a Biosemi Active Two 32-channel EEG system. In accordance with the international 10-20 system, the EEG was recorded from 32 sites localized around the sensorimotor area. The selected electrodes are $FC_5, FC_3, FC_1, FC_z, FC_2, FC_4, FC_6, C_5, C_3, C_1, C_z, C_2, C_4, C_6, CP_5, CP_3, CP_1, CP_z, CP_2, CP_4, CP_6, P_3, P_1, P_z, P_2, P_4, PO_3, PO_z, PO_4, O_1, O_z, O_2$. An external electromyogram (EMG) electrode was used in order to verify that there was no movement during the KMI task.

D. Methods

1) *Riemannian geometry for BCI*: We use the Riemannian geometric framework described in [6] to encode the statistical features of experimental EEG recordings via their spatial covariance matrices. These matrices are generally symmetric positive definite (SPD) and all data operations are done respecting the intrinsic geometry of the SPD manifold. In the classification tasks described next, we use the minimum distance to mean classifier (MDM) [6], which is a generalization of the classical nearest-centroid classifier to data points in the SPD manifold.

2) *Transfer learning*: Our transfer learning problem concerns the statistical mismatch between two BCI datasets defined in the SPD manifold: a source dataset (\mathcal{S}) and a target dataset (\mathcal{T}). This mismatch explains why training a classifier on \mathcal{S} and applying it directly to \mathcal{T} often yields poor performance. In what follows, we consider a semi-supervised setting for BCI transfer learning, where the labels of all trials of \mathcal{S} are known in advance and only the first labels from \mathcal{T} are available, forming a “target learning” subset $\mathcal{T}_L \subset \mathcal{T}$.

We use the Riemannian Procrustes Analysis (RPA) [8] to match the statistics of \mathcal{S} and \mathcal{T} . This method considers the distributions of the data points as shapes in a high-dimensional space and performs rigid geometric operations, such as re-centering, stretching and rotation, to make the data shapes as similar as possible. The RPA transformations are done respecting the intrinsic geometry of the SPD manifold

and using the supervised information available in \mathcal{S} and \mathcal{T}_L . Note that RPA is expected to work better when more trials are available in \mathcal{T}_L , since it leads to better estimates of the statistics of the target dataset. We refer the reader to [8] for further details on the definitions and inner workings of RPA.

We considered four classification pipelines when analysing the data from our BCI experiments. They differ in how the training ($\mathcal{D}_{\text{train}}$) and testing ($\mathcal{D}_{\text{test}}$) sets for the MDM classifier are defined:

- DCT The source and target datasets are used Directly as training and testing datasets, without any transformation, i.e. $\mathcal{D}_{\text{train}} = \mathcal{S}$ and $\mathcal{D}_{\text{test}} = \mathcal{T} \setminus \mathcal{T}_L$.
- RCT The source and target datasets are Re-CenTered to the origin of the SPD manifold using the first step of the RPA algorithm, which is completely unsupervised in (the source and) target set, i.e. it does not use any information from the labels of \mathcal{T}_L . We have $\mathcal{D}_{\text{train}} = \mathcal{S}^{(\text{RCT})} \cup \mathcal{T}^{(\text{RCT})}$ and $\mathcal{D}_{\text{test}} = \mathcal{T}^{(\text{RCT})} \setminus \mathcal{T}_L^{(\text{RCT})}$.
- RPA We apply the full RPA algorithm to the source and target datasets, getting $\mathcal{D}_{\text{train}} = \mathcal{S}^{(\text{RPA})} \cup \mathcal{T}_L^{(\text{RPA})}$ and $\mathcal{D}_{\text{test}} = \mathcal{T}^{(\text{RPA})} \setminus \mathcal{T}_L^{(\text{RPA})}$.
- CLB This is the usual CaLiBration, where the classifier is trained using only the trials available in \mathcal{T}_L . We have $\mathcal{D}_{\text{train}} = \mathcal{T}_L$ and $\mathcal{D}_{\text{test}} = \mathcal{T} \setminus \mathcal{T}_L$.

Note that the classification performances that we report for each pipeline were not cross-validated, since we consider a realistic setting where the data from the trials become available sequentially.

3) Statistical tests:

a) *Two-way ANOVA with repeated measures*: is used to study statistical differences between means of accuracies over subjects according to two factors: pipeline (DCT, RCT, RPA, CLB) and number of labeled trials in the target dataset (2, 4, 24 and 40) – this subset ensures a wide enough range of analysis while avoiding too many statistical tests. If needed, we adjusted the data sphericity using Greenhouse–Geisser’s correction.

b) *Paired t-test*: detects statistical differences between the average performance of each pipeline across subjects, for different number of trials. We used the paired t-tests as post-hoc tests, to identify pair-wise differences following the ANOVA described above. Since the number of trials for all subjects are identical for all approaches the test is paired. We adjusted the p -values for the multiple comparisons problem using Holm’s step-down method using Bonferroni adjustments.

III. RESULTS

A. Comparing transfer learning and standard approaches

Figure 1 showed that using only the labeled samples of the target dataset to train a classifier (CLB) is less efficient than using also labeled samples of the source dataset (DCT, RCT or RPA), especially when the number of labeled samples is low. For all approaches the accuracy increases with the number of labeled samples of the target dataset. This is

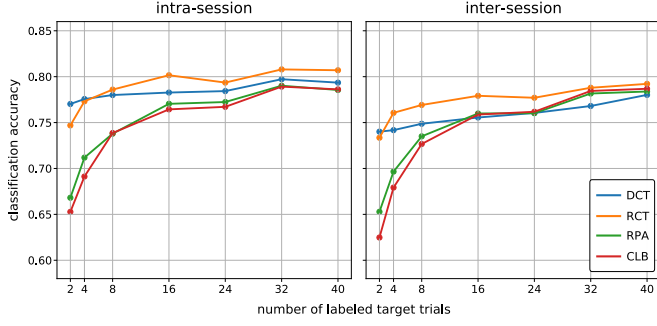


Fig. 1: Mean accuracies over 30 subjects according to the number of labeled trials in the target dataset for four different learning approaches DCT, RCT, RPA, CLB (see section II-D.2) for intra- and inter-session transfer learning.

particularly visible for CLB and RPA. Two-ways ANOVA shown statistical difference ($p\text{-GG-corr} < 0.001$) between approaches and number of labeled trials in the target group for intra-session and inter-session learnings. Paired t-tests show the following significant differences:

- ($p\text{-corr} < 0.001$): $\text{RCT} > \text{RPA}$, $\text{RCT} > \text{CLB}$, $\text{DCT} > \text{RPA}$, $\text{DCT} > \text{CLB}$ for intra- and inter-session learning.
- ($p\text{-corr} < 0.01$): $\text{RPA} > \text{CLB}$ for intra-session learning.
- ($p\text{-corr} < 0.05$): $\text{RPA} > \text{CLB}$ for inter-session learning.

B. Comparing intra-session and inter-session evaluations

Statistical differences obtained comparing accuracies of methods by t-tests are the same for intra-session and for inter-session transfer learning (see subfigures on Figure 1) except comparing RPA and CLB (see section III-A). So, methods and in particular transfer learning ones do not obtain better results with inter-session evaluations than with intra-session evaluations, on our dataset.

C. Effect of the accuracy of the source dataset

To investigate if the transfer learning has a different effect according to the accuracy of the source dataset, the population of subjects has been splitted in two identically sized groups namely the inferior and superior groups according to their accuracy on the source session vis-a-vis the median accuracy. Figure 2 shows that the recentering seems even more useful for subjects with a low accuracy on the source dataset for inter-session transfer learning (see section IV-B for an interpretation).

IV. DISCUSSION

A. Is transfer learning better than not doing anything?

Results shown that RCT presents better results than other methods when using a few target trials (see also section IV-C). The centering operator can be rapidly estimated. RPA needs more trials especially to estimate the rotation operator since the stretching is usually almost constant with EEG. CLB starts with only one trial to build the classifier and even with more labeled target data shows lower accuracies. Then merging labeled samples of the source and target dataset

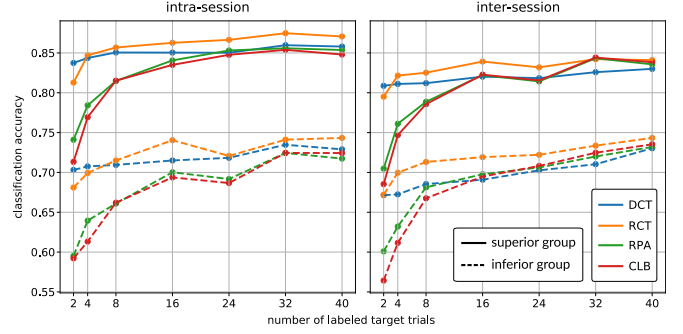


Fig. 2: Mean accuracies over 30 subjects according to the percentage of trials in the target dataset for four different learning approaches DCT, RCT, RPA, CLB (see section II-D.2). Solid lines: subjects with an accuracy above the median accuracy of the calibration session, dashed lines: subjects with an accuracy below the median accuracy of the calibration session.

(DCT) is a good solution, and even better if samples are recentered (RCT). However merging rotated samples requires more labeled samples of the target dataset.

B. Who can benefit from transfer learning?

Section III-C showed a higher benefit for subjects with a lower accuracy on the source data set (namely the inferior group). We hypothesize that subjects with a lower accuracy present higher difficulties to produce kinesthetic motor imagery which is an unusual and non-tangible task. Their EEG trials and so their covariance matrices could be more diverse, thus requiring a stronger alignment on data of the source session.

Figure 3 shows that a subject with a high accuracy presents stable features even between sessions while a subject with a low accuracy (i.e with mixed class distributions) has very different features from one session to another and will gain more benefits from transfer learning.

C. How many target trials do we need for an effective transfer learning?

First, figures shown that RCT outperforms other methods using 8 target labeled trials on our dataset. For inter-session transfer learning, fewer target labeled trials seem necessary. One explanation could be that if the distributions of the source and target data are more different, which should be the case for inter-session target data as compared to intra-session one, then a simple merge of the source and target data would be less effective than a recentering of the target data. Secondly, RPA and CLB reach DCT performances from 16 target trials and outperform them from 32 trials. Nevertheless, they do not outperform RCT and when the goal is to reduce the calibration time more than 32 trials require around 5 min of use.

V. CONCLUSION

In this article, we confirmed the value of using data from a previous session to reduce the calibration time of the session

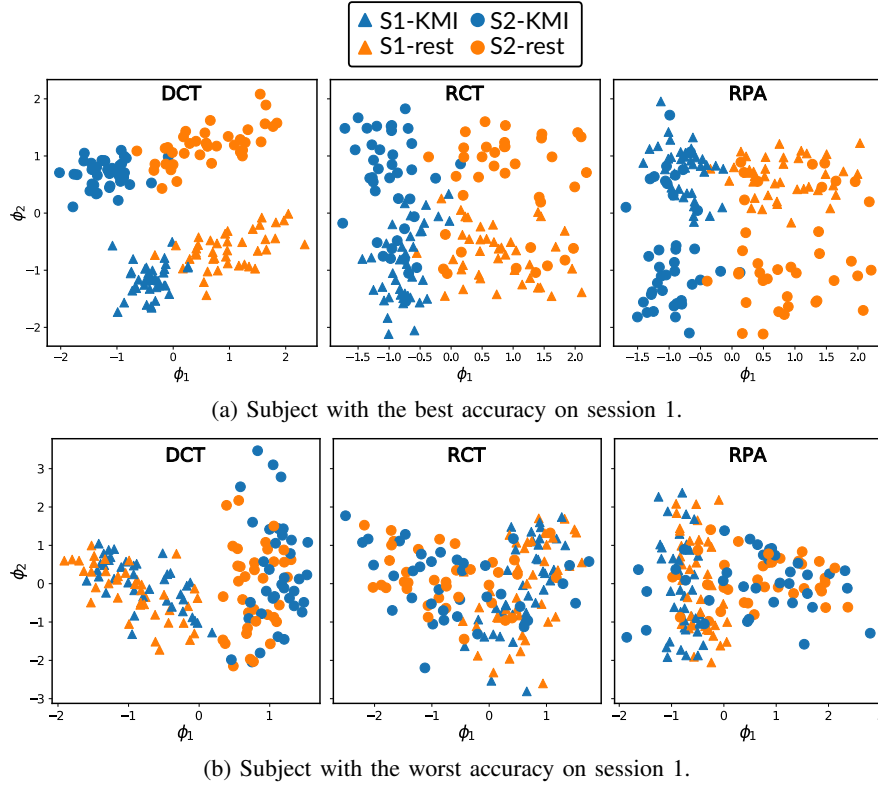


Fig. 3: Covariance matrices (i.e. features) of subjects with the best (a) and worst (b) accuracy on the source dataset are projected in a 2D-space using diffusion maps [11]. S1 and S2 refer to the entire dataset of sessions 1 and 2 respectively. Features are presented without geometric operations (DCT), after a re-centering (RCT), and after a re-centering, stretching and rotation (RPA), to make the data shapes as similar as possible.

of the day and increase the performance of brain-computer interfaces based on motor imagery. More precisely, we have shown that the unsupervised transfer learning operation of recentring the data (RCT) increases the accuracy from a few target trials. Subjects having difficulty producing motor imagery benefit even more from this operation. In contrast, the supervised operation of rotating to align the target data with the source data (RPA) requires more trials to be beneficial. As part of a reduction in calibration time, this operation brings fewer benefits.

ACKNOWLEDGMENT

The author acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-19-CE33-0007 (Grasp-IT project), and the European Research Council with project BrainConquest (grant ERC-2016-STG-714567).

REFERENCES

- [1] M. Clerc, L. Bougrain, and F. Lotte, *Brain-Computer Interfaces 1: Foundations and Methods*, M. Clerc, L. Bougrain, and F. Lotte, Eds. John Wiley & Sons, 2016.
- [2] S. Rimbart, P. Riff, N. Gayraud, D. Schmartz, and L. Bougrain, "Median nerve stimulation based bci: A new approach to detect intraoperative awareness during general anesthesia," *Frontiers in Neuroscience*, vol. 13, p. 622, 2019.
- [3] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for eeg-based brain-computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.
- [4] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, June 2015.
- [5] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: A review," *IEEE Transactions on Neural System and Rehabilitation Engineering*, vol. PP, 10 2017.
- [6] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01570120>
- [7] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: A riemannian geometry framework with applications to brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1107–1116, 2017.
- [8] P. L.C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, 2018.
- [9] P. L.C. Rodrigues, M. Congedo, and C. Jutten, "“When does it work?”: An exploratory analysis of transfer learning for BCI," in *Int. Graz BCI Conf.*, 2019.
- [10] S. Rimbart, L. Bougrain, and S. Fleck, "Learning How to Generate Kinesthetic Motor Imagery Using a BCI-based Learning Environment: a Comparative Study Based on Guided or Trial-and-Error Approaches," in *Systems Man and Cybernetics 2020*, 2020.
- [11] S. Lafon and A. B. Lee, "Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1393–1403, 2006.